Fostering Compositionality in Generative RNNs to Solve the Omniglot challenge

Sarah Fabi, Sebastian Otte, & Martin V. Butz
Neuro-Cognitive Modeling, University of Tübingen

Introduction: To motivate researchers to investigate how efficient learning based on compositionality, causality, and learning to learn can be realized within machine learning algorithms, the Omniglot challenge (Lake et al., 2015) was developed. It consists of several one- and few-shot classification and generation tasks of handwritten character trajectories. Previous models that attempted to solve the challenge did provide motor primitives, focused on just some of the tasks or solved tasks only inadequately (Lake et al., 2019). We wanted to investigate whether we could solve the challenge with a simple generative model without providing motor primitives but by fostering and reusing compositional encodings.

Model and One-shot inference mechanism: We applied a generative LSTM model with a linear embedding layer (100 neurons) and a recurrent layer (100 LSTM units), which, at every timestep, got one-hot vectors encoding the character as input and predicted the change in x and y position. We trained the network on 440 trajectories of the first half of the Latin alphabet of sequential data of handwritten characters. For the one-shot tasks, when presented with one variant of a new character, we allowed only the first weights into the embedding layer to adapt. If representations of components had been developed during training, this mechanism should re-arrange them to form new characters.

Results: The different tasks of the challenge were met with the help of our mechanism and a newly recorded sequential Latin alphabet: One-shot regeneration, classification, generation of new variants and of totally new character concepts. To investigate whether compositionality led to the success of our model, we analysed the LSTM cell and hidden states more thoroughly with the t-distributed stochastic neighbour embedding (t-SNE), which projected the hidden state activations onto a 2d-space (cf. Figure 1). The cell states c were clearly clustered per character (example: half circle + downwards stroke for ‘q’ and ‘y’) and might serve as an indicator for the network to stay in the correct attractor. For the hidden states h, clear character components could be identified (example: bottom to top trajectories on the left). This speaks for the hypothesis that the network extracted compositional encodings during training that it could later reuse to efficiently learn new characters. In further experiments, we trained the network on alphabets of the original Omniglot dataset and let it solve the one-shot tasks on other alphabets. If the alphabets shared components, this worked well, whereas if the components that were necessary for the one-shot tasks were not part of the training alphabet, the results looked worse (cf. Figure 2), speaking again for our hypothesis that compositionality helps solving the challenge.

Conclusion: By fostering compositionality in latent, generative encodings, the Omniglot challenge was met without providing knowledge about motor primitives (as done in Lake et al., 2015). During training on some characters, the model learns components that it can recombine, when it is confronted with new characters. The t-SNE analysis provides evidence that such compositional structures develop in the hidden states of the LSTM cells. This research is a step towards bringing Machine Learning algorithms towards closer resemblance to human cognitive mechanisms by fostering compositionality in generative LSTM networks.

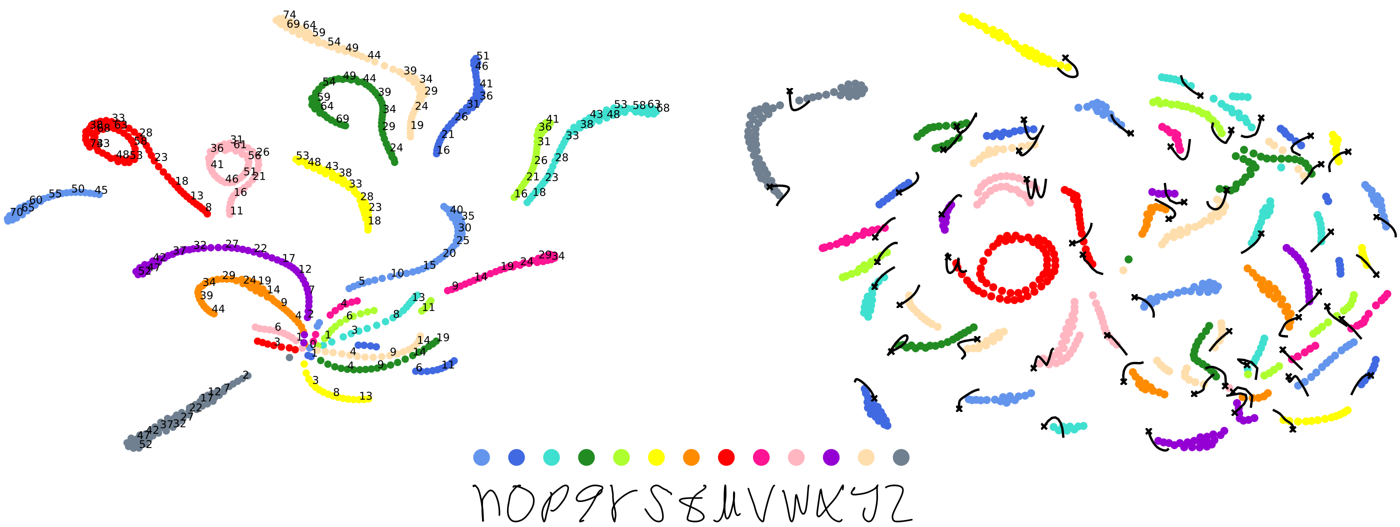
Additional Materials

<u>References</u>

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*, 1332-1338.

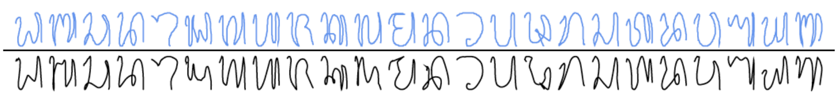
Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2019). The Omniglot challenge: A 3-year progress report. *Current Opinion in Behavioral Sciences*, *29*, 97-104.

<u>Figure1: t-SNE 2-dimensional representations of cell states c (left) and hidden states h (right)</u>

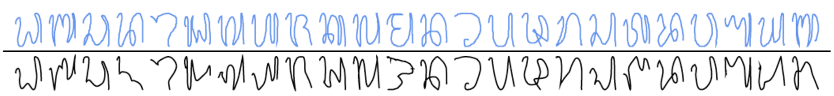


<u>Figure 2: One-shot regeneration task on Balinese alphabet, trained on alphabets with different amounts of overlapping components</u>

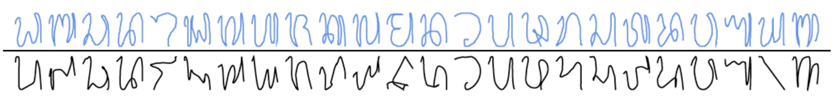
(a) Training on first half of the Balinese alphabet provided all necessary components leading to best regeneration.



(b) Training on the Burmese alphabet provided most necessary components leading to good regeneration.



(c) Training on the Latin alphabet did not provide all necessary components leading to worse regeneration.



(d) Training on the Greek alphabet provided even less components leading to worst regeneration.